DOCUMENT RESUME

ED 247 298 TM 840 479

AUTHOR Loyd, Brenda H.
TITLE Efficiency and Precision in Two-Stage Adaptive Testing.
PUB DATE Feb 84
NOTE 13p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (West Palm Beach, FL, February 9-11, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Ability; *Adaptive Testing; *Difficulty Level; Elementary Education; Latent Trait Theory; *Quantitative Tests; Test Items; Test Reliability

ABSTRACT

One form of adaptive testing involves a two-stage procedure. The first stage is the administration of a routing test. From this first test, an estimate of an examinee's ability is obtained. On the basis of this ability estimate, a second test focused on a given ability level is administered. The purpose of this study was to compare the efficiency of three routing tests of different lengths (10, 15, and 20 items). The routing tests and one of six 40-item second-stage tests measuring mathematics concepts, were administered to 1439 students, grades 3-8. The three routing tests were compared on the basis of the consistency with which they assigned examinees to the same second-stage test, and in terms of the accuracy with which they indicated the most appropriate second-stage test. The increase in information attained by employing a routing test (with cut-off values based on the raw score scale) over that which would result from assigning examinees to test levels on the basis of grade level alone suggest that more precise measurement may be possible with a two-stage procedure than with a single stage. The 20-item routing test was most effective. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED247298

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EFFICIENCY AND PRECISION IN TWO-STAGE ADAPTIVE TESTING

Brenda H. Loyd

University of Virginia

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
B. H. Loyd
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the Annual Meeting of the
Eastern Educational Research Association,
West Palm Beach, February, 1984

TM 840 479

EFFICIENCY AND PRECISION IN TWO-STAGE ADAPTIVE TESTING

Brenda H. Loyd

University of Virginia

Conventional achievement and ability measures are constructed by determining a best "fixed set" of items for a group of examinees with a specific level of ability. Such a test usually measures very precisely for the levels at which the test is focused or "peaked." This is usually accomplished by selecting a majority of the items with difficulty values (p-values) around .5 for the group specified. However, the difficulty of the test relative to a specific individual will vary. For the individual for whom the test is too difficult or too easy, the test will not provide optimal measurement. A test with a wider range of item difficulties, while providing equal measurement across a wode range of ability levels, will provide lower levels of precision at any specific level of ability.

In response to this problem adaptive testing methods based on item response theory have been proposed (see, for example, Weiss, 1982). Adaptive testing methods hold the promise of permitting measurements of equal precision throughout the range of a trait while maintaining high levels of efficiency. Adaptive testing administered by computer programs could involve measuring each individual with a different set of items which would optimize measurement for that individual.

One form of adaptive testing involves a two-stage procedure and may be administered by paper and pencil methods as well as by computers. The first stage is the administration of a routing test. From this first test, an estimate of an examinee's ability is obtained. On the basis of this ability estimate, a second test is administered. This second test is a set of items focused or peaked at a given ability level.

The purpose of this study was to compare the efficiency of three routing tests with different numbers of items. The three routing tests were compared on the basis of the consistency with which they assigned examinees to the same second

stage test. The routing tests were also compared in terms of the accuracy with which they indicated the most appropriate second-stage test.

Methodology

The trait of interest was mathematics concepts, which includes understanding of mathematical processes and systems. Three first-stage or routing tests were constructed, containing respectively 10 items, 15 items, and 20 items. Each routing test was composed of different items, but each set of items was selected to have approximately the same average difficulty. Each set of items also included a wide range of difficulty appropriate for students in grades three through eight.

For the second stage, a series of six tests were used, composed of from 35 to 44 items. The series of measures were considered peaked at appropriate levels for each of the six grade groups from grade three through grade eight.

The subjects were 1439 students, with approximately 250 students from each of the grades from third through eight.

Each student was administered each of the three routing tests and one of the six 40-item second-stage tests. However, for the purposes of the study, the second-stage tests were assigned according to grade, not according to the results of the routing tests.

The item responses were analyzed by using a modified three-parameter logistic model. Estimates of item parameters and examinee abilities were calculated using the LOGIST computer program (Wood, Wingersky, and Lord, 1976). The c-parameters were estimated using the entire sample and then were fixed at an average value. With the c-parameter fixed, estimates of the difficulty and discrimination parameters (b and a) were obtained for the routing tests separately using the entire sample, and for each of the six second-stage tests separately using the subsamples for each grade group.

The decision as to which of the second-stage tests should be given depends on the ability estimate obtained on the first-stage or routing test. Estimates

of ability were made on two bases. The first was on raw score. Cut off points on the raw score scales of each of the three routing tests were set to maximize the separation among the six grade groups. For example on the 10-item routing test, examinees who received a raw score of zero, one, or two would be assigned the lowest level second-stage test (the level appropriate for grade 3), examinees who received a raw score of 3 or 4 would be assigned the second lowest level second-stage test (the level appropriate for grade 4). The cut-off values for each of the three routing tests are present in Table 1. The second basis was according to a weighted raw score where the item responses were weighted by the item discrimination parameters. Again cut off points were set to maximize differences among the grade groups. The cut-off points on the weighted raw score scale are also presented in Table 1.

INSERT TABLE 1 ABOUT HERE.

After estimates of each examinee's theta (ability parameter in the logistic item characteristic curve model) were calculated and item parameters for all six second stage tests had been estimated, it was possible to estimate the value of the information function for each of the second stage tests for each examinee. Therefore, six information values were obtained for each examinee. For each of the three routing tests and for each of the two methods of assignment, the information corresponding to the second-stage test to which the examinee was assigned could be determined. Across all examinees the best routing test and routing method would be the one which assigned the examinees to the second-stage test which provided the most information about the examinee. Therefore, the mean information obtained as a result of each assignment procedure was obtained. To determine the consistency among the routing tests and the methods of assignment, correlations were also computed.

Results

The means for each routing test by each scoring method are presented in Table 2 for each of the six grade groups and for the total sample. The means of

the information values for each of the six second stage tests and for each of the routing test procedures are presented in Table 3. For comparison, the mean information value when the second stage test was assigned only on the basis of grade was also included in Table 3.

The correlations among the information values for each of the assigned procedures is presented in Table 4. The correlations among second-stage tests designated by each of the assignment procedures are also presented in Table 4.

INSERT TABLES 2, 3 AND 4 ABOUT HERE

The correlations among the six methods of assignment indicated relatively high agreement on the second stage test level assigned (see Table 4). The correlations ranged from .75 to .95. Correlations between the method based on grade and each of the six methods based on routing test were somewhat lower in agreement on which level was assigned. The correlations ranged from .66 to .72.

The correlations among the six methods indicated moderate agreement for information values but less than that obtained for assignment to levels. These correlations as presented in Table 4 ranged from .48 to .85. The correlations between assignment by grade and each of the six methods of assignment based on routing tests for information values ranged from .44 to .55 and indicated lower agreement relative to the other sets of correlations.

An examination of the average information values presented in Table 3 indicates that the largest average information resulted from the assignment based on the twenty item routing test and the raw score scale. The average information for this method was 8.27. The assignment method which used the fifteen item routing test and the raw score cut off produced the second highest average information, a value of 8.16. All three routing tests which employed the raw score scaled values resulted in average information values greater than that obtained by assigning by grade designation alone.

The three methods which employed the weighted scores were lower in average information (ranging from 7.22 to 7.71) than the methods based only on raw score and were lower than the average information that resulted by assignment according to grade (7.89). The average information did increase with an increase in the number of items in the routing test.

Discussion

The results of the study provide some support for the use of adaptive testing to increase the efficiency and precision in a two-stage testing situation. The increase in information by employing a routing test (with cut-off values based on the raw score scale) over that which would result from assigning examinees to test levels on the basis of grade level alone suggest that more precise measurement may be possible with a two stage procedure than with a single stage.

The results of the study also suggest that assignments based on a weighted raw score, obtained by weighting correct item responses by the item discrimination parameter estimate, did not produce more precise assignment of second-stage test level and, in fact, resulted in less precise measurement than that provided by grade assignment alone. This clearly does not imply that all weighted scores would be less effective. The use of other weighted score scales would need to be investigated.

The use of a twenty-item routing test was more effective than the fifteen or ten-item routing test and the fifteen-item test was more effective than the ten-item test in assigning examinees to the second-stage test level that would increase the information in their measurement. Although, the differences in information obtained by increasing the number of items in the routing test would be statistically significant because of the large sample size, the practical implications of the difference in information obtained by using longer routing tests would have to be examined before one could state that longer tests are more effective.

The testing situation which is explored in this study is consistent with the elementary school achievement testing situation. In such testing, levels of achievement tests are developed to measure a part of the developmental continuum more efficiently. A given test level may focus on a single grade and contain items which measure the typical student in that grade level most efficiently. However, for the students who are atypical, either advanced or slow in their development, the level designed for their grade group may be inappropriate for them. For those advanced in their development, such a level would provide minimal challenge and for those who are slow in their development, such a level may be frustrating and discouraging. In addition, for these atypical students, the precision or reliability of the resulting measurement is suspect. A two stage testing procedure in which each student is given a short routing test and then, based on the scores on the routing test, the most appropriate test level is given, may have the advantage of increasing the precision of the measurement as well as proving the attitudes of the examinees towards testing.

Table 1

Routing Test Values for

Assigning Second-Stage Test

Method	Second-Stage Test					
	Level 1 (Grade 3)	Level 2 (Grade 4)	Level 3 (Grade 5)	Level 4 (Grade 6)	Level 5 (Grade 6)	Level 6 (Grade 8)
10-item Raw Score	0-2	3, 4	5	6	7	8-10
10-item Weighted Score	0-2.9	2.9-4.0	4.0-5.1	5.1-6.0	6.0-8.0	8.0-10.0
15-item Raw Score	0-3	4-5	6-7	8	9-10	11-15
15-item Weighted Score	0-4.5	4.5-6.6	6.6-8.3	8.3-9.7	9.7-10.9	10.9-15
20-item Raw Score	0-5	6-8	9-10	11-12	13-14	15-20
20-item Weighted Score	0-6.8	6.8-8.9	8.9-10.8	10.8-12.8	19.8-14.5	14.5-20

Table 2

Means of Raw Scores and

Weighted Raw Scores for Each Grade Group and Total Sample

Method	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Total
10-item Raw Score	2.47	3.65	4.57	5.78	6.90	7.98	5.16
10-item Weighted Score	2.29	3.48	4.48	5.59	6.41	7.20	4.85
15-item Raw Score	3.12	5.22	6.86	8.01	9.53	11.03	7.21
15-item Weighted Score	3.20	5.72	7.63	9.01	10.39	11.40	7.80
20-item Raw Score	5.36	7.66	9.18	11.28	13.61	15.25	10.26
20-item Weighted Score	5.34	8.10	9.72	11.74	13.82	15.08	10.51
	N=250	N=247	N=259	N=218	N=225	N=236	N=1435

Table 3

Means and Standard Deviations of

Information for Each Method

of Assigning Second-Stage Test

Method	Information	
	Mean	SD
10-item Raw Score	7.98	3.57
10-item Weighted Score	7.22	3.55
15-item Raw Score	8.16	3.69
15-item Weighted Score	7.52	3.67
20-item Raw Score	8.27	3.64
20-item Weighted Score	7.71	3.64
Assigned by Grade	7.89	3.59
All Assigned Level 1 (Grade 3)	3.16	1.43
All Assigned Level 2 (Grade 4)	8.19	3.92
All Assigned Level 3 (Grade 5)	9.82	4.05
All Assigned Level 4 (Grade 6)	7.66	2.96
All Assigned Level 5 (Grade 7)	5.34	2.60
All Assigned Level 6 (Grade 8)	3.55	1.34

Table 4

Correlations of Information Values and of

Second-stage Assignments Decisions Among Assigning Procedures

Method	Correlations for Information Values (Upper Right) and for Levels Assigned (Lower Left)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
10-item Raw Score (1)	---	.60	.65	.79	.54	.62	.54
10-item Weighted Score (2)	.76	---	.66	.52	.84	.63	.51
15-item Raw Score (3)	.80	.83	---	.56	.62	.85	.55
15-item Weighted Score (4)	.94	.75	.78	---	.48	.53	.44
20-item Raw Score (5)	.75	.95	.82	.75	---	.59	.46
20-item Weighted Score (6)	.80	.84	.95	.79	.84	---	.53
Assignment by Grade	.71	.69	.72	.67	.66	.71	---

References

Lord, F.M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum, 1980

Weiss, D.J. Improving measurement quality and efficiency with adaptive testing. Applied Psychological Measurement, 1982, 6, 473-492.